Final Analysis Paper

Patrick McGrath, James Yang, Kristen Dolan

April 27, 2021

Introduction

The goal of the electroencephalogram (EEG) research capstone project is to determine whether low-cost equipment can be used with data processing and filtering techniques to visualize and predict hand movement. Research in this area is being conducted with two primary objectives in mind: better understanding how the brain and electrical signals control the body, and developing technology that can mimic human functionality. Given that this system will infer hand motion from brain signals, this is an example of using neural signals to decode human intent. Past the scope of this project, a practical example of this technology being utilized would be to allow amputees to control their prosthetics with brain signals. This project's focus on low-cost equipment contributes to the possibility of expanding the accessibility of the technology to encourage equity and reduce discrimination.

Broadly, this system will be composed of a data collection block, a data processing block, a machine learning block, and a data visualization block. The team's goal by the end of this project is to have established a data collection system to store data simultaneously, designed preprocessing systems and feature selection methods that will remove artifacts and identify hand motion, developed an interactive data visualization tool for the user, trained a machine learning model with customizable hyperparameters, and displayed the live prediction of hand movement from the system virtually for the user to view. The data collection portion of the system is what the low-cost technology directly affects, and the team plans to use methods of preprocessing and feature selection that can account for and eliminate the unwanted artifacts in the collected data.

The team has researched, independently implemented, and tested three combinations of preprocessing and feature selection methods that are well documented in existing literature. The goal is to determine their functionality with static, verified sets of EEG data before applying them to the newly gathered data to narrow the opportunity for unidentifiable error. The next steps to be taken are to develop the data collection system and data visualization tool to begin analysis on the quality of data produced by the low-cost equipment.

This document serves as a detailed description of my area of expertise within the realm of the team's chosen preprocessing and feature selection methods for our project partner, the corresponding research group, and other parties involved in the project. It will detail the application and uses for common average referencing, linear discriminant analysis, spatial pattern filtering, specifically common spatial patterns, principal component analysis, independent component analysis, and will discuss accuracies that our research has presented in comparison to each other's methods.

Equipment Used

EEG tests will primarily come from headsets. In our research project, we will be using a 16-channel "Ultracortex Mark IV EEG Headset". The data that we collect will be noisy, especially since the readings come in millivolts. This headset will be connected to an interface that is proprietary to analyzing EEG data. The headset is capable of sampling rates between 990-1000hz. It is also capable of gathering substantial amounts of data with a margin of error of 5%. We are also using a proprietary glove that can collect data with a sampling rate of 33-34hz.

Figure 1: Depicts the BCI headset from OPENBCI shop.

All of this is being outputted through a software called "Mujoco", a hand visualization software that can represent what is occurring with the headset and glove. The software is integrated within the interface designed by the Oregon State Information Processing Group and can properly output current and predicted hand movement.

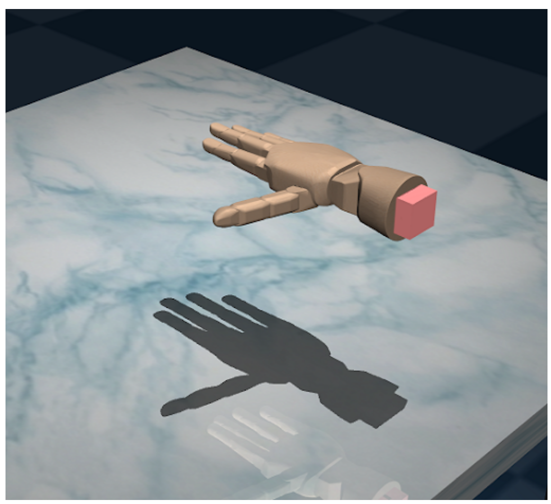

Figure 1.1: Depicts the MuJoco hand visualization.

Feature Generating Methods

K. Liao, R. Xiao, J. Gonzalez, L. Ding Method (ICA and PSD Method)

All gathered data from this experiment was from 11 healthy right-handed participants. A high pass filter was applied to the EEG data at .3Hz and a notch filter was applied to eliminate power line noise at 60Hz. ICA was also applied to the data to examine EEG specific data without the additional noise. Finger movement was band-pass filtered between 0.5-2Hz to eliminate outside noise. Power spectral density is then applied to the data to analyze particular frequency bands. Principal component analysis is then applied to the PSD algorithm to highlight the most volatile features. Shown in **figure 1.1** is the waves in which the algorithm distributes its values from. A five-fold cross validation was applied to the data using an SVM machine learning model. For this research paper, the authors employed PSD and PCA onto EEG data before feeding the newly generated features into an SVM [1]. With the high accuracy attained in this paper of **77.11%,** utilizing an SVM for classification is a good benchmark.

$$P_n(f,\tau_q) = \frac{1}{T}\left|\sum_{t=-\frac{T}{2}}^{\frac{T}{2}-1} X_n(\tau_q + t)\cdot H(t)\cdot \exp\left(i\frac{2\pi}{T}(f-1)t\right)\right|^2$$

$$f = 1,2,\cdots,N_f \quad \tau_q = 1,2,\cdots,N_q$$

Figure 1.2: Depicts the PSD wave distribution equation.

E. Neto, F. Biessmann, H. Aurlien, H. Nordby, T. Eichele (CAR and LDA Method)

LDA assigns observations to corresponding classes based on a set of measurements or predictors by finding an optimal linear transformation that maximizes class separability. It achieves optimal solutions by using predictor vectors (multivariate) that are normally distributed within each class and different groups that have similar covariance. Models can be overfitted and predictability overestimated which is why they aren't going to be 90 and above type accuracy. Noise can be captured from adjacent channels if they are close enough. The solution to these issues could be applying spatial filters and regularization. Regularization with cross validation replaces covariance by a weighted average of the whole sample. This will increase larger eigenvalues while decreasing smaller ones which creates a pooled-covariance matrix. Shown in **figure 1.2**, they achieved an accuracy of **83.0%** with this method.

	Complete set of features		Reduced set of features	
Classifier performances	**cv-ACC**	**AUC**	**cv-ACC**	**AUC**
Model 1 (HC vs. AD)	0.62	0.66	0.67	0.74
Model 2 (HC vs. VaD)	0.65	0.68	0.72	0.77
Model 3 (AD vs. VaD)	0.59	0.62	0.57	0.61
Model 4 (HC vs. AD&VaD)	0.70	0.75	0.77	0.83

Average of the CV accuracy (cv-ACC) and AUC performed for each final classifier using two different sets of features. The Complete Set of Features included a total of 132 predictors and the Reduced Set of Features included variable number of predictors based on automatic feature reduction. Values from [0.5–0.6] = Fail; [0.6–0.7] = Poor; [0.7–0.8] = Fair; [0.8–0.9] = Good; [0.9–1.0] = Excellent.

Figure 1.3: Depicts the RLDA accuracy.

H. Lu, H. Eng, C. Guan, K. Plataniotis, A. Venestansopoulos (CSP Method)

This paper proposed and tested a combination of parameter regularization and aggregation that served to eliminate the issue of having a small sample size of data. Their algorithm, shown in **figure 1.3**, integrates a variety of regularization parameters and the use of the nearest neighbor classifier into the classic CSP algorithm to increase the performance. Using the Regularized Common Spatial Pattern with Aggregation (R-CSP-A) algorithm proposed in the article by Lu et al will benefit our final product by allowing for less error in feature classification, requiring less initial training time and data, and being unique to each user's EEG signal patterns [3]. Once fully implemented the performance of this algorithm will be compared using a static, verified data set and the new data collected using our system to analyze the benefit it brings to our application of low-cost equipment. It is clear that this algorithm is high performing in this setting and will lead to better performance in a majority of the trial cases. According to their paper, they have generated for R-CSP-A **83.4%** accuracy [3].

Input: A set of M EEG trials $\mathbf{E}_{(c,m)_{(s)}}$ for each class of S subjects, where $c = \{1, 2\}$, $m = 1, 2, ..., M$, and $s = 1, ..., S$. A test trial $\mathbf{E}$ for subject s^*, A pairs of β and γ, the number of most discriminative columns from the full projection matrix $Q = 2\alpha$. Subject s^* is considered as the subject of interest and other subjects with $s \neq s^*$ are considered as the generic data.
Output: The class label for $\mathbf{E}$.
R-CSP-A algorithm:

Step 1. Feature extraction

- Obtain $\mathbf{S}_{(c,m)}$ for all subjects $s = 1, ..., S$ according to (1).
- Form $\mathbf{S}_c$ for subject s^* according to (5) and form $\hat{\mathbf{S}}_c$ from other subjects $s \neq s^*$ according to (6).
- For $a = 1 : A$
  - Follow (4), (3), (7), and (8) to get the full projection matrix.
  - Retain the first and last α columns of the full projection matrix to get $\hat{\mathbf{W}}_{(a)}$.
  - Follow (9) and (10) to obtain the feature vector $\hat{\mathbf{y}}_{(a)}$.

Step 2. Aggregation at the matching score level for classification

- For $a = 1 : A$
  - Apply (11) and (12) on $\hat{\mathbf{y}}_{(a)}$ to get $\mathbf{z}_{(a)}$.
  - For $c = 1 : 2$
    - Obtain the nearest-neighbor distance $d(\mathbf{E}, c, a)$.
  - Normalize $d(\mathbf{E}, c, a)$ to $[0, 1]$ to get $\tilde{d}(\mathbf{E}, c, a)$.
- Obtain the aggregated distance $d(\mathbf{E}, c)$.
- **Output** $c^* = \arg\min_c d(\mathbf{E}, c)$ as the class label for the test sample $\mathbf{E}$.

Figure 1.4: Depicts the Lu Algorithm.

Testing/Accuracy

PSD Method of Testing

Three pairs of comparisons were performed between left hand movement, right hand movement, and no movement for each of the subjects. A support vector machine was trained for each of these comparisons with a radial basis kernel function. This allowed the data to be compared and separated at a high dimensional space. Gamma and the penalty parameter were determined from a grid-based search [4]. A two-layer multi layer perceptron network was used to classify between left hand movement, right hand movement, and no movement. One hidden layer was used with a relu activation function. The output layer had a soft-max activation function to differentiate between the different classes. The Adam optimizer was used for the neural network. We achieve accuracy results of **74.3%** and **74.8%** for classifying movement versus no movement using a support vector machine and multi layer perceptron network, respectively.

LDA Method of Testing

After running data through common average referencing and linear discriminant analysis, it appears using a binary decoder that the training accuracy comes out to be **94.8%** and **52.7%** on testing accuracy with 4310 voltage readings of training and 1000 voltage readings of testing. The decoder is using an array of prediction that gets values of the labels of hand movement. If the voltage of the label is greater than 0, then it is considered open handed. Otherwise, if the label is 0, then it is considered closed handed. To calculate the accuracy, we simply sum all of the values and divide them by the total number of elements (labels.shape[0]). This gives us a percentage of what was correct and what was incorrect.

```
prediction[prediction >= 0] = 1
prediction[prediction < 0] = -1
accuracy = ((prediction == labels).sum()) / labels.shape[0]
```

Figure 1.5: Depicts the decoder algorithm used for LDA.

CSP Method of Testing

In CSP, there are decoding methods of normalizing band-power features from 9 arbitrary chosen frequencies between 0-33hz. They translate into trajectory predictions by the weight of the matrix obtained from the linear regression. We can apply these types of predictions on the data collected from the headset, and utilize them by training and then testing the data. The output from the CSP accuracy pattern was inconsistent and ranged from **60-68%** within testing.

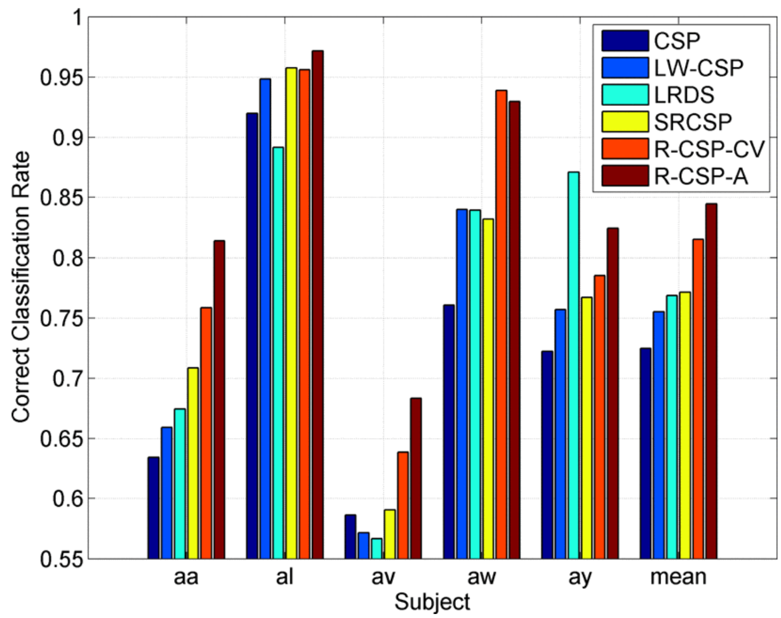

Figure 1.6: Depicts the CSP algorithm performance comparison from a researched article.

Conclusion

To conclude, all algorithms depict different results of accuracy using completely different methodologies. After implementing the three methods, we have come to the conclusion that ICA and PSD have had the highest testing accuracy. With 74.8% accuracy, we can safely conclude that it passes the threshold of meeting 60% for the senior capstone engineering requirement. It also properly satisfies a successful feature generating method paired with a pre-processing method.

References

[1]. K. Liao, R. Xiao, J. Gonzalez, and L. Ding, "Decoding individual finger movements from one hand using human EEG signals," PloS one, 08-Jan-2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3885680/. [Accessed: 21-Jan-2021].

[2]. Neto E, Biessmann F, Aurlien H, Nordby H, Eichele T. Regularized Linear Discriminant Analysis of EEG Features in Dementia Patients. Front Aging Neurosci. 2016;8:273. Published 2016 Nov 30. doi:10.3389/fnagi.2016.00273

[3]. H. Lu, H. Eng, C. Guan, K. N. Plataniotis and A. N. Venetsanopoulos, "Regularized Common Spatial Pattern With Aggregation for EEG Classification in Small-Sample Setting," in IEEE Transactions on Biomedical Engineering, vol. 57, no. 12, pp. 2936-2946, Dec. 2010, doi: 10.1109/TBME.2010.2082540. https://ieeexplore-ieee-org.ezproxy.proxy.library.oregonstate.edu/document/5593203 [Accessed March 3, 2021]

[4]. Hsu, Chih-Wei, Chih-Chung Chang, and Chih-Jen Lin. "A practical guide to support vector classification. Department of Computer Science, National Taiwan University." (2010).